Sample Midterm

7:00-9:00pm, 7 March

Read these instructions carefully

1. This is a **closed book** exam, but you are allowed one single-sided cheat sheet. No phones, calculators or other electronic equipment.
2. The exam consists of 10 questions. The first seven questions are multiple choice; the remaining three require written answers. *Note: The actual midterm will have a similar format but may have different numbers of questions of the two types.*
3. Approximate point totals for each question part are indicated in the margin. The maximum total number of points is 92.
4. **Multiple choice questions:** Answer these by **circling** the correct answer. You should be able to answer all of these from memory, by inspection, or with a very small calculation. Incorrect answers will receive a **negative** score, so if you do not know the answer you should **not** guess. There is no partial credit for these.
5. **Other questions:** Write your answers to these in the spaces provided below them. None of these questions requires a long answer, so you should have enough space; if not, continue on the back of the page and state clearly that you have done so. **Show your working**.
6. The questions vary in difficulty: if you get stuck on some part of a question, leave it and go on to the next one.

Your Name:

---

1. You are dealt a hand of five cards from a randomly shuffled deck.

(a) The probability that your hand contains no pair of cards with the same numerical value (there are 13 numerical values: ace, 2, . . ., king) is *3pts*

$\left(\frac{12}{13}\right)^4$ $\quad 1-\left(\frac{12}{13}\right)^4$ $\quad \frac{48\cdot47\cdot46\cdot45}{51\cdot50\cdot49\cdot48}$ $\quad \frac{48\cdot44\cdot40\cdot36}{51\cdot50\cdot49\cdot48}$ $\quad \frac{13\cdot12\cdot11\cdot10\cdot9}{52\cdot51\cdot50\cdot49\cdot48}$ $\quad 4\cdot\frac{12\cdot11\cdot10\cdot9}{51\cdot50\cdot49\cdot48}$

(b) The probability that the hand contains *at least three* cards of the same value is *3pts*

$\frac{13(4\cdot\binom{48}{2}+48)}{\binom{52}{5}}$ $\quad \frac{13\cdot4\cdot48\cdot47}{\binom{52}{5}}$ $\quad \frac{13\cdot48}{\binom{52}{5}}$ $\quad \left(\frac{1}{13}\right)^3$ $\quad 5\cdot\frac{3\cdot2}{51\cdot50}$ $\quad \frac{13\cdot\binom{5}{3}\cdot\binom{4}{3}}{\binom{52}{5}}$

---

2. A coin with heads probability p is tossed n times independently. The random variable X measures the number of heads observed *minus* the number of tails observed.

(a) The *expectation* of X is *3pts*

0 $\quad np$ $\quad n(p-1)$ $\quad n(2p-1)$ $\quad np(1-p)$

(b) The *variance* of X is *3pts*

np $\quad np^2$ $\quad n^2p^2$ $\quad 2np(1-p)$ $\quad 4np(1-p)$

[continued overleaf]

3. We have $2n$ bins and we throw $2n$ balls into them in two phases as follows. In the first phase, the first n balls are thrown into the $2n$ bins, independently and uniformly at random. In the second phase, the remaining n balls are thrown into only the first n bins, independently and uniformly at random.

(a) After both phases, the expected number of balls in the first *three* bins is *3pts*

$1 \qquad \frac{3}{2} \qquad 3 \qquad \frac{9}{2} \qquad \frac{n}{3} \qquad n$

(b) After both phases, when n is large the expected number of empty bins among the first three bins is about *3pts*

$3e^{-3/2} \qquad e^{-3/2} \qquad 3e^{-3} \qquad e^{-3} \qquad 3e^{-2} \qquad e^{-2}$

---

4. Two fair six-sided dice are rolled. The random variables X_1, X_2 denote the score on the first die and the second die respectively, and $X = X_1 + X_2$ is the total score on both dice.

(a) The conditional expectation $E[X|X_1 = 3]$ is *3pts*

$3 \qquad \frac{7}{2} \qquad 6 \qquad \frac{13}{2} \qquad 7$ $\qquad$ none of these

(b) The conditional expectation $E[E[X|X_1]]$ is *3pts*

$3 \qquad \frac{7}{2} \qquad 6 \qquad \frac{13}{2} \qquad 7$ $\qquad$ none of these

---

5. Recall the coupon-collecting problem, in which each cereal box contains one of n different coupons, each coupon being equally likely. Here are two variants of the standard question.

(a) The expected number of boxes that need to be bought until $n/2$ different coupons are obtained is on the order of *3pts*

$\log n \qquad n \qquad n \log\log n \qquad n \log n \qquad n^2 \qquad 2^n$

(b) The expected number of boxes that need to be bought until the number of coupons of which no copy has been obtained is less than $\sqrt{n}$ is on the order of *3pts*

$\log n \qquad n \qquad n \log\log n \qquad n \log n \qquad n^2 \qquad 2^n$

---

6. A fair six-sided die is rolled repeatedly until *four* different outcomes are observed. The expected number of times the die is tossed is *3pts*

$\frac{6}{5} \qquad \frac{11}{4} \qquad 3 \qquad \frac{33}{5} \qquad \frac{57}{10} \qquad 6$

---

7. Let X and Y be *independent* random variables on the same probability space, with $E[X] = 3$, $\mathrm{Var}[X] = 3$, and Y a 0-1 random variable with $E[Y] = \frac{2}{3}$. Circle those <u>three</u> of the following statements that <u>must</u> be true about X and Y: *6pts*

$\Pr[X \geq 6] \leq \frac{1}{3} \qquad \Pr[X \geq Y] < 1 \qquad E[X^2] = 12$

$\mathrm{Cov}(X, Y) = \frac{2}{3} \qquad \mathrm{Var}[Y^2] = \frac{2}{9} \qquad \mathrm{Var}[X + 3Y] = \frac{11}{3}$

[continued overleaf]

8. Sampling With and Without Replacement

A bag contains 10 green and 10 yellow balls. Two samples, of $n \leq 20$ balls each, are taken from the bag. The first sample is with replacement, and the second sample is without replacement. Let the random variables X and Y denote the number of green balls in the first and second sample respectively.

(a) What is $E[X]$ (as a function of n)? *3pts*

(b) What is $\text{Var}[X]$ (as a function of n)? *3pts*

(c) Let us write $Y = \sum_{i=1}^{n} Y_i$, where $Y_i = 1$ if the ith ball in the sample without replacement is green, and $Y_i = 0$ otherwise. Explain briefly why Y_i has the same distribution as Y_1 for all i. *2pts*

(d) Using part (c), compute $E[Y]$ (as a function of n). *3pts*

[continued overleaf]

(e) Compute $\mathrm{Var}[Y]$ (as a function of n). [HINT: Follow a similar path to your computation of $E[Y]$ in parts (c) and (d). You might want to check your answer by substituting in the special values $n = 1$ and $n = 20$.] *6pts*

(f) Comment briefly on the difference between $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$. *2pts*

[continued overleaf]

9. Coloring a Graph

A *3-coloring* of an undirected graph $G = (V, E)$ is an assignment of colors red, green or blue to every vertex in the graph. An edge is *well-colored* if its two endpoints are assigned different colors. In the problem MAX3COLOR, we are given an undirected graph $G = (V, E)$ and asked to find a 3-coloring with the maximum possible number of well-colored edges. MAX3COLOR is an NP-hard problem, so we do not expect to find an efficient algorithm that solves it exactly. Here is a very simple linear-time randomized algorithm that gives a pretty good approximation:

(1) Randomly and independently color each vertex $v \in V$ red, green or blue with probability $\frac{1}{3}$ each

(2) Output the resulting 3-coloring

Let the r.v. X denote the number of well-colored edges in the 3-coloring output by the algorithm. In addition, for every edge $e \in E$, let X_e be the indicator r.v. that assumes the value 1 if e is well-colored and 0 otherwise.

(a) Write down the equation relating the random variable X and the random variables X_e. *2pts*

(b) Show that $E[X] = \frac{2}{3}m$ where $m = |E|$. Deduce that $E[X] \geq \frac{2}{3}\text{OPT}$, where OPT is the maximum number of well-colored edges over all possible 3-colorings of G. *4pts*

(c) Compute $E[X_e X_{e'}]$ for any $e \neq e' \in E$. *3pts*

[continued overleaf]

(d) Compute $\mathrm{Var}[X]$. *3pts*

(e) Let p denote the probability that the 3-coloring output by the algorithm has at least $\frac{5}{9}\text{OPT}$ well-colored edges. By applying Chebyshev's inequality to the r.v. X, show that $p = 1 - O(1/m)$. *5pts*

(f) What is wrong with the following argument, which claims to give a better bound on p: *2pts*

> *X is the sum of 0-1-r.v.'s. Applying the Chernoff bound to X, in the form $\Pr[X \le (1-\delta)\mu] \le e^{-\delta^2\mu/2}$, with $\mu = \frac{2m}{3}$ and $\delta = \frac{1}{6}$, we may deduce that $\Pr[X \le \frac{5}{9}\text{OPT}] \le e^{-\frac{1}{72}\cdot\frac{2m}{3}} = 2^{-\Omega(m)}$. Hence, $p = 1 - 2^{-\Omega(m)}$.*

[continued overleaf]

10. Estimating a Parameter

Recall from class the following procedure for estimating the fraction p of Republicans in a given city:

(1) Take a random sample of t people (with replacement), and set $X_i = 1$ if the ith person is Republican, and 0 otherwise

(2) Output $X = \frac{1}{t}\sum_{i=1}^{t} X_i$

Our goal is to obtain an estimate that is within relative error ϵ with confidence $1-\delta$, i.e., we want to achieve

$$\Pr[|X - p| \geq \epsilon p] \leq \delta. \qquad (*)$$

where ϵ and δ are parameters. (Again, we have seen this in class.)

(a) Use a Chernoff bound to show that, in order to achieve the goal in equation $(*)$, it is enough to take $t = \frac{c\ln(2/\delta)}{\epsilon^2 p}$, for some constant c. [NOTE: We have seen this in class. Do not worry about the exact value of c.] *5pts*

(b) We now look at a different mechanism for ensuring the goal in equation $(*)$. This mechanism involves a two-stage procedure: *2pts*

(1) Run the above procedure s times, with the number of trials t on each run chosen to satisfy equation $(*)$ for the fixed value $\delta = 1/4$

(2) Output the *median* of the s values obtained

Show that the required value of t in step (1) is at most $\frac{c'}{\epsilon^2 p}$, for some constant c'. [Again, don't worry about the exact value of c'.]

[continued overleaf]

(c) Show that, in order to achieve the goal in equation $(*)$, it is enough to take $s = c'' \ln(1/\delta)$ for a constant c''. [HINT: Use a Chernoff bound applied to a coin with a suitable heads probability. Don't worry about the exact value of c''.] *6pts*

(d) Deduce from parts (b) and (c) that the total number of samples needed in the new scheme is $c''' \frac{\ln(1/\delta)}{\epsilon^2 p}$ for some constant c''' (and thus of the same order of magnitude as proved for the original scheme in part (a)). *2pts*

[The End]